

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-049617

(43)Date of publication of application : 15.02.2002

(51)Int.Cl.

G06F 17/27

(21)Application number : 2000-232576

(71)Applicant : SAN FUREA:KK

(22)Date of filing : 01.08.2000

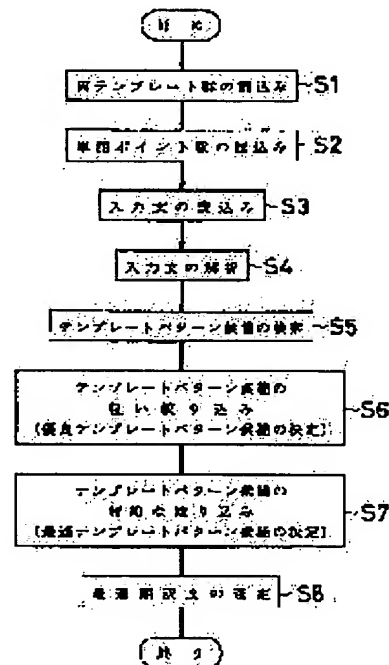
(72)Inventor : TOKUDA NAOYUKI
CHIN AKIRA
SASAI HIROYUKI

(54) METHOD AND DEVICE FOR SELECTION OF BEST TRANSLATION AND RECORDING MEDIUM

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a method and a device for selection of best translation and recording medium capable of preparing the best translation of high precision in a short time.

SOLUTION: Template group for original texts (English for example) and template group for translation (Japanese for example) are prepared. Both templates mutually related, when English sentences are selected in a template group for original texts, Japanese sentences corresponding to this English sentences are kept to be gotten in the translation group. All words in the template group for original texts are put with point of point numbers according to the order of importance. And both template groups and points are read in (S1, S2). After that input sentences are read in and are analyzed (S2, S3). Then candidates for template pattern are searched (S5), the rough narrowing down (S6) and meticulous narrowing down (S7) are practiced sequentially, the best candidate for template pattern is decided. And a Japanese candidate for template pattern corresponding to it is selected as the best translation (S8). As the candidate for template pattern is narrowed down by two steps, a user can get the best meticulous translation in a short time.



LEGAL STATUS

[Date of request for examination]

13.06.2003

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

3778785

[Date of registration]

10.03.2006

[Number of appeal against examiner's decision
of rejection]

[Date of requesting appeal against examiner's
decision of rejection]

[Date of extinction of right]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2002-49617

(P2002-49617A)

(43) 公開日 平成14年2月15日 (2002.2.15)

(51) IntCl.⁷

G 0 6 F 17/27

識別記号

F I

G 0 6 F 17/27

データベース (参考)

J 5 B 0 9 1

審査請求 未請求 請求項の数 3 O L (全 21 頁)

(21) 出願番号 特願2000-232576 (P2000-232576)

(22) 出願日 平成12年8月1日 (2000.8.1)

(71) 出願人 596091794

株式会社サン・フレア

東京都新宿区四谷4丁目7番地

(72) 発明者 徳田 尚之

東京都国分寺市富士本2丁目29番3号

(72) 発明者 陳 亮

栃木県宇都宮市峰町253番地1 醍醐ハイ
ツ A102号

(72) 発明者 笹井 紘幸

東京都福生市加美平3丁目35番10号

(74) 代理人 100080078

弁理士 駒津 敏洋 (外1名)

Fターム (参考) 5B091 AA01 BA03 CA21 CC01 CC15

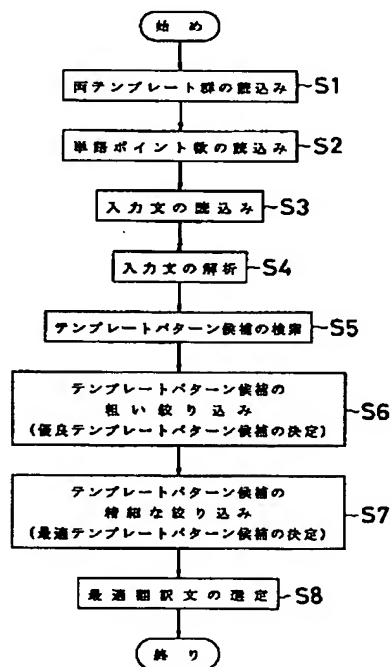
EA24

(54) 【発明の名称】 最適翻訳文選定方法、選定装置および記録媒体

(57) 【要約】

【課題】 精度の高い最適翻訳文を、短時間で得られるようにする。

【解決手段】 原文用のテンプレート群 (例えば英語) と、訳文用テンプレート群 (例えば日本語) とを用意する。両テンプレート群を相互に関連付けておき、原文用のテンプレート群から英語の文章を選んだならば、これに対応する日本語の文章が、訳文用テンプレート群から得られるようにしておく。原文用のテンプレート群中のすべての単語に、その重要度に応じたポイント数のポイントを付けておく。そして、両テンプレート群およびポイントを読み込む (S1, S2)。その後入力文を読み込むとともに、その解析を行なう (S2, S3)。次いで、テンプレートパターン候補を検索し (S5)、その粗い絞り込み (S6) および精細な絞り込み (S7) を順次行ない、最適テンプレートパターン候補を決定する。そして、これに対応する日本語のテンプレートパターン候補を、最適翻訳文として選定する (S8)。テンプレートパターン候補を二段で絞り込んでいるので、精度の高い最適翻訳文を、短時間で得ることができる。



【特許請求の範囲】

【請求項1】 A T N機能を有する任意数の文章要素を同列で包含する複数のテンプレートブロックで構成される原文用のテンプレート群およびこのテンプレート群と意味的に等価な訳文用テンプレート群が、相互に関連付けて記憶されている翻訳用データベースを用い、前記テンプレート群から作成可能なすべてのテンプレートパターン候補の中から、入力文に最も適したテンプレートパターン候補を検索するとともに、このテンプレートパターン候補に対応する訳文用テンプレートパターン候補を、前記訳文用テンプレート群から作成し、この訳文用テンプレートパターン候補を、最適翻訳文として選定する最適翻訳文選定方法であって、前記テンプレート群中のすべての単語に、その重要度に応じたポイント数のポイントを付与した後、前記すべてのテンプレートパターン候補および入力文について、各テンプレートパターン候補および入力文で用いられているすべての単語のテンプレート群中における出現頻度とその単語のポイント数とに基づき、各テンプレートパターン候補および入力文の正規化L S Iベクトルを求めるとともに、S V Dアルゴリズムを用いて設定される圧縮文書L S Iベクトル空間に射影し、入力文に対し圧縮文書L S Iベクトル空間における類似度が高い順に複数のテンプレートパターン候補を選んで優良テンプレートパターン候補とし、次いで各優良テンプレートパターン候補で用いられているすべての単語の合計ポイント数を分母、入力文と各優良テンプレートパターン候補との間での一致単語の合計ポイント数を分子とする最適比較値を求め、分子が最大となる最適比較値のうちで、最適比較値が最も大きくなる優良テンプレートパターン候補を選ぶとともに、これに対応する訳文用テンプレートパターン候補を、最適翻訳文として選定することを特徴とする最適翻訳文選定方法。

【請求項2】 A T N機能を有する任意数の文章要素を同列で包含する複数のテンプレートブロックで構成される原文用のテンプレート群およびこのテンプレート群と意味的に等価な訳文用テンプレート群とを、相互に関連付けて記憶する翻訳用データベースと；入力文を記憶する入力文記憶手段と；前記テンプレート群の各テンプレートブロックと入力文とを比較し、テンプレート群から作成可能なすべてのテンプレートパターン候補を検索するテンプレートパターン候補検索手段；検索されたテンプレートパターン候補を記憶するテンプレートパターン候補記憶手段と；テンプレート群中のすべての単語にその重要度に応じたポイント数で付与されるポイントを、単語と関連付けて記憶する単語ポイント数記憶手段と；各テンプレートパターン候補および入力文で用いられるすべての単語のテンプレート群中における出現頻度を、各テンプレートパターン候補および入力文毎に単語と関連付けて記憶する単語出現頻度記憶手段と；単語の出現

頻度とポイント数とに基づき、各テンプレートパターン候補および入力文の正規化L S Iベクトルを演算する正規化L S Iベクトル演算手段と；演算された各テンプレートパターン候補および入力文の正規化L S Iベクトルを記憶する正規化L S Iベクトル記憶手段と；S V Dアルゴリズムを用いて設定される圧縮文書L S Iベクトル空間において、各テンプレートパターン候補の正規化L S Iベクトルと入力文の正規化L S Iベクトルとを比較し、その類似度を演算する類似度演算手段と；類似度が高い方から順に選ばれた複数のテンプレートパターン候補を、優良テンプレートパターン候補として記憶する優良テンプレートパターン候補記憶手段と；各優良テンプレートパターン候補において入力文と一致する単語を検索する一致単語検索手段と；各優良テンプレートパターン候補で用いられているすべての単語の合計ポイント数を演算する合計ポイント数演算手段と；入力文と各優良テンプレートパターン候補との間での一致単語の合計ポイント数を演算する一致単語ポイント数演算手段と；各優良テンプレートパターン候補における全単語の合計ポイント数を分母、各優良テンプレートパターン候補における入力文との一致単語の合計ポイント数を分子とする最適比較値を演算する最適比較値演算手段と；各優良テンプレートパターン候補の最適比較値を比較し、分子が最大となる最適比較値のうちで、最も大きな最適比較値を有する優良テンプレートパターン候補を、最適テンプレートパターン候補として決定する最適テンプレートパターン候補決定手段と；最適テンプレートパターン候補に対応する訳文用テンプレートパターン候補を、最適翻訳文として選定する最適翻訳文選定手段と；を具備することを特徴とする最適翻訳文選定装置。

【請求項3】 A T N機能を有する任意数の文章要素を同列で包含する複数のテンプレートブロックで構成される原文用のテンプレート群およびこのテンプレート群と意味的に等価な訳文用テンプレート群を、相互に関連付けてメモリ領域に記憶させる処理；入力文をメモリ領域に記憶させる処理；前記テンプレート群中のすべての単語に対し、その重要度に応じたポイント数で付与されたポイントを、単語と関連付けてメモリ領域に記憶させる処理；前記テンプレート群から作成可能なすべてのテンプレートパターン候補および入力文について、各テンプレートパターン候補および入力文で用いられているすべての単語のテンプレート群中における出現頻度とその単語のポイント数とに基づき、各テンプレートパターン候補および入力文の正規化L S Iベクトルを求め処理；S V Dアルゴリズムを用いて設定される圧縮文書L S Iベクトル空間において、各テンプレートパターン候補の正規化L S Iベクトルと入力文の正規化L S Iベクトルとを比較し、その類似度を求める処理；類似度の高い方から順に選んだ複数のテンプレートパターン候補を、優良テンプレートパターン候補としてメモリ領域に記憶さ

せる処理：各優良テンプレートパターン候補で用いられるすべての単語の合計ポイント数を分母、入力文と各優良テンプレートパターン候補との間での一致単語の合計ポイント数を分子とする最適比較値を用い、分子が最大となる最適比較値のうちで、最適比較値が最も大きくなる優良テンプレートパターン候補を、最適テンプレートパターン候補として決定する処理；最適テンプレートパターン候補に対応する訳文用テンプレートパターン候補を、最適訳文として選定する処理；をコンピュータに実行させるためのプログラムを記録したコンピュータ読取り可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、最適訳文選定方法、選定装置および記録媒体に係り、特に既存のコンピュータ・プログラム用マニュアルの内容を一部改訂して改訂版を作る等の際に用いるのに好適な最適訳文選定方法、選定装置および記録媒体に関する。

【0002】

【従来の技術】一般に、コンピュータ・プログラムは全世界で販売されることが多いため、そのマニュアルは、英語を初めとして、日本語、フランス語、ドイツ語、中国語等、多数の言語を用いて作成する必要がある。

【0003】従来、このような場合には、まず基体となる言語、例えば英語でマニュアルを作成し、この英語に基づき、各言語の専門家が日本語等の各言語に翻訳する方法が採られている。そしてこの場合、例えば日本語マニュアルの各文章は、英語マニュアルの各文章と意味的に完全に等価なものとなっている。

【0004】

【発明が解決しようとする課題】ところで、コンピュータ・プログラムは、その内容を部分的に改訂して新たな機能を付加することがしばしば行なわれるが、部分的に改訂した部分については、マニュアルの内容も訂正、変更あるいは追加することが必要となる。

【0005】従来、このような部分的な改訂も人力で行なっているが、例えば英語マニュアルの改訂部分が日本語マニュアルのどの部分に該当し、その内容がどのように変更されたのかを理解するためには、日本語の知識のみならず、英語の知識も必要となり、そのような翻訳専門家を確保することは容易でないという問題がある。

【0006】そこで、改訂部分については、コンピュータを用いて機械翻訳することが考えられるが、従来の機械翻訳では、単語単位で意味を当て嵌め、これを当該言語の文法規則に基づき並べ替える等の単純な方法でしか翻訳を行っていないため、実際に使用できるような訳文が得られないという問題がある。

【0007】ところで、前記コンピュータ・プログラム用マニュアルのような場合、改訂部分以外の既存部分については、各言語間において、各文章が完全に対一で

対応している。したがって、改訂部分について、各文章が既存部分ではどのように翻訳されているかを検討し、既存部分における同一文章あるいは類似文章の訳文を、改訂部分の文章の訳文として当て嵌めることにより、極めて短時間で、しかも極めて精度の高い訳文が得られるものと予想される。

【0008】本発明は、かかる知見に基づきなされたもので、機械翻訳であっても、極めて短時間で精度の高い訳文を得ることができる最適訳文選定方法、選定装置および記録媒体を提供することを目的とする。

【0009】

【課題を解決するための手段】前記目的を達成するため本発明は、A T N機能を有する任意数の文章要素を同列で包含する複数のテンプレートブロックで構成される原文用のテンプレート群およびこのテンプレート群と意味的に等価な訳文用テンプレート群が、相互に関連付けて記憶されている翻訳用データベースを用い、前記テンプレート群から作成可能なすべてのテンプレートパターン候補の中から、入力文に最も適したテンプレートパターン候補を検索するとともに、このテンプレートパターン候補に対応する訳文用テンプレートパターン候補を、前記訳文用テンプレート群から作成し、この訳文用テンプレートパターン候補を、最適訳文として選定する最適訳文選定方法であって、前記テンプレート群中のすべての単語に、その重要度に応じたポイント数のポイントを付与した後、前記すべてのテンプレートパターン候補および入力文について、各テンプレートパターン候補および入力文で用いられているすべての単語のテンプレート群中における出現頻度とその単語のポイント数とに基づき、各テンプレートパターン候補および入力文の正規化L S Iベクトルを求めるとともに、SVDアルゴリズムを用いて設定される圧縮文書L S Iベクトル空間に射影し、入力文に対し圧縮文書L S Iベクトル空間における類似度が高い順に複数のテンプレートパターン候補を選んで優良テンプレートパターン候補とし、次いで各優良テンプレートパターン候補で用いられているすべての単語の合計ポイント数を分母、入力文と各優良テンプレートパターン候補との間での一致単語の合計ポイント数を分子とする最適比較値を求め、分子が最大となる最適比較値のうちで、最適比較値が最も大きくなる優良テンプレートパターン候補を選ぶとともに、これに対応する訳文用テンプレートパターン候補を、最適訳文として選定するようにしたことを特徴とする。

【0010】すなわち、本発明においては、例えば英語マニュアルの一部が改訂され、この改訂された部分を、日本語に翻訳するような場合、まず英語マニュアルの既存部分と日本語マニュアルの既存部分とが、相互に関係付けられたテンプレート群の形式で翻訳用データベースに記憶される。したがって、翻訳用データベース内においては、英語のテンプレート群（原文用のテンプレート

群)から得られる文章と、日本語のテンプレート群(訳文用テンプレート群)から得られる文章とが、一対一で対応していることになる。

【0011】この状態で、英語の入力文(改訂部分の英語の文章)が入力されると、この入力文と同一あるいは最も類似する文章が、英語のテンプレートパターンの形式で翻訳用データベースから選ばれ、これに対応する日本語のテンプレートパターンが、最適翻訳文として選定されることになる。

【0012】ここで、翻訳用データベースに記憶されているデータ量が多ければ多い程、最適翻訳文の選定精度が向上することは明らかであるが、反面データ量が多くなれば、最適翻訳文の選定処理に多くの時間を要し実用に供し得ない。

【0013】ところが本発明においては、正規化LSIベクトルを圧縮文書LSIベクトル空間に射影する方法を用い、粗い選定を行なって対象となるテンプレートパターンの数を絞り、その後絞られた少数のテンプレートパターンに対し、最適比較値を用いた詳細選定を行なうようにしている。このため、精度の高い最適翻訳文を短時間で得ることが可能となる。

【0014】本発明はまた、ATN機能を有する任意数の文章要素を同列で包含する複数のテンプレートブロックで構成される原文用のテンプレート群およびこのテンプレート群と意味的に等価な訳文用テンプレート群とを、相互に関連付けて記憶する翻訳用データベースと；入力文を記憶する入力文記憶手段と；前記テンプレート群の各テンプレートブロックと入力文とを比較し、テンプレート群から作成可能なすべてのテンプレートパターン候補を検索するテンプレートパターン候補検索手段；検索されたテンプレートパターン候補を記憶するテンプレートパターン候補記憶手段と；テンプレート群中のすべての単語にその重要度に応じたポイント数で付与されるポイントを、単語と関連付けて記憶する単語ポイント数記憶手段と；各テンプレートパターン候補および入力文で用いられるすべての単語のテンプレート群中における出現頻度を、各テンプレートパターン候補および入力文毎に単語と関連付けて記憶する単語出現頻度記憶手段と；単語の出現頻度とポイント数とに基づき、各テンプレートパターン候補および入力文の正規化LSIベクトルを演算する正規化LSIベクトル演算手段と；演算された各テンプレートパターン候補および入力文の正規化LSIベクトルを記憶する正規化LSIベクトル記憶手段と；SVDアルゴリズムを用いて設定される圧縮文書LSIベクトル空間において、各テンプレートパターン候補の正規化LSIベクトルと入力文の正規化LSIベクトルとを比較し、その類似度を演算する類似度演算手段と；類似度が高い方から順に選ばれた複数のテンプレートパターン候補を、優良テンプレートパターン候補として記憶する優良テンプレートパターン候補記憶手段

と；各優良テンプレートパターン候補において入力文と一致する単語を検索する一致単語検索手段と；各優良テンプレートパターン候補で用いられているすべての単語の合計ポイント数を演算する合計ポイント数演算手段と；入力文と各優良テンプレートパターン候補との間での一致単語の合計ポイント数を演算する一致単語ポイント数演算手段と；各優良テンプレートパターン候補における全単語の合計ポイント数を分母、各優良テンプレートパターン候補における入力文との一致単語の合計ポイント数を分子とする最適比較値を演算する最適比較値演算手段と；各優良テンプレートパターン候補の最適比較値を比較し、分子が最大となる最適比較値のうちに、最も大きな最適比較値を有する優良テンプレートパターン候補を、最適テンプレートパターン候補として決定する最適テンプレートパターン候補決定手段と；最適テンプレートパターン候補に対応する訳文用テンプレートパターン候補を、最適翻訳文として選定する最適翻訳文選定手段と；を設けるようにしたことを特徴とする。そして、類似度演算手段等を用い、対象となるテンプレートパターンの数を優良テンプレートパターン候補として絞った後、最適テンプレートパターン候補決定手段等を用い最適なテンプレートパターン候補を決定するようにしているので、短時間で精度の高い最適翻訳文を選定することが可能となる。

【0015】本発明はさらに、ATN機能を有する任意数の文章要素を同列で包含する複数のテンプレートブロックで構成される原文用のテンプレート群およびこのテンプレート群と意味的に等価な訳文用テンプレート群を、相互に関連付けてメモリ領域に記憶させる処理；入力文をメモリ領域に記憶させる処理；前記テンプレート群中のすべての単語に対し、その重要度に応じたポイント数で付与されたポイントを、単語と関連付けてメモリ領域に記憶させる処理；前記テンプレート群から作成可能なすべてのテンプレートパターン候補および入力文について、各テンプレートパターン候補および入力文で用いられているすべての単語のテンプレート群中における出現頻度とその単語のポイント数とに基づき、各テンプレートパターン候補および入力文の正規化LSIベクトルを求める処理；SVDアルゴリズムを用いて設定される圧縮文書LSIベクトル空間において、各テンプレートパターン候補の正規化LSIベクトルと入力文の正規化LSIベクトルとを比較し、その類似度を求める処理；類似度の高い方から順に選んだ複数のテンプレートパターン候補を、優良テンプレートパターン候補としてメモリ領域に記憶させる処理；各優良テンプレートパターン候補で用いられるすべての単語の合計ポイント数を分母、入力文と各優良テンプレートパターン候補との間での一致単語の合計ポイント数を分子とする最適比較値を用い、分子が最大となる最適比較値のうちに、最適比較値が最も大きくなる優良テンプレートパターン

候補を、最適テンプレートパターン候補として決定する処理；最適テンプレートパターン候補に対応する訳文用テンプレートパターン候補を、最適翻訳文として選定する処理；をコンピュータに実行させるようにしたことを特徴とする。そして前記各処理により、精度の高い最適翻訳文を短時間で得ることが可能となる。

【0016】

【発明の実施の形態】以下、本発明を図面を参照して説明する。図1は、本発明の実施の一形態に係る最適翻訳文選定装置を示すもので、この装置は、例えば英語を原文とする場合の原文用のテンプレート群およびこのテンプレート群と意味的に等価な例えば日本語の訳文用テンプレート群が相互に関連付けて記憶されている翻訳文データベース1、例えば英語の入力文を記憶する入力文記憶手段2、およびテンプレートパターン候補検索手段3を備えており、テンプレートパターン候補検索手段3は、後に詳述する原文用のテンプレート群の各テンプレートブロックと入力文とを比較し、原文用のテンプレート群から作成可能なすべてのテンプレートパターン候補を探索するようになっている。そして、探索された各テンプレートパターン候補は、テンプレートパターン候補記憶手段4に記憶されるようになっている。

【0017】前記原文用のテンプレート群は、図2に示すように、単語・句からなる文章要素を任意数同列で包含する複数のテンプレートブロックを備えており、これら各テンプレートブロックは、文法規則に基づき上位のものから順次配列されている。そして、このテンプレート群からは、上位のテンプレートブロック内の任意の文章要素と、その下位のテンプレートブロック内の任意の文章要素とを順次連結していくことにより、入力文に対応する多数のテンプレートパターン候補を作成することができるようになっている。

【0018】例えば、図2に示すテンプレート群の場合には、

『Japan has beautiful park snationwide.』

『Japan is dotted with lovely public gardens all over the country.』

等のテンプレートパターン候補を作成することができるようになっている。

【0019】このテンプレート群は、過去のマニュアル作成の際の添削例等を参考にして、添削作業等で翻訳文の核になる要素が何であるかを、専門家の知識を用いて吟味し、既存のマニュアルの文章を単語・句からなる文章要素に分解するとともに、文法規則上同順位の文章要素を任意数同列でまとめてテンプレートブロックを作り、各テンプレートブロックを、文法規則に基づき上位のものから順次配列することにより構成されている。

【0020】前記各テンプレートブロックは、正しく使

われた単語・句だけでなく、統語的または意味上誤用された単語・句も含めて作成されており、これにより、通常考えられないような入力文に対しても、これに対応するテンプレートパターン候補を作成することができるようになっている。

【0021】また、前記テンプレート群中のすべての単語には、図2に示すように、その重要度に応じたポイント数のポイントが付与されており、このポイントは、図1に示す単語ポイント数記憶手段5に、単語と関連付けて記憶されるようになっている。

【0022】例えば、図2に示すテンプレート群の場合には、ほとんどの単語には、ポイント数「1」のポイントが付与されているが、『dotted』および『there』には、他の単語の2倍のポイント数「2」のポイントが付与されているとともに、『nationwide』には、他の単語の3倍のポイント数「3」のポイントが付与されている。すなわち、『dotted』、『there』および『nationwide』は、他の単語の2倍のあるいは3倍の重要度を有する重要単語として設定されている。

【0023】この重要単語は、例えばマニュアルを作成するような場合に、特定の動作には特定の用語を用いるのが通例であり、したがって、マニュアルの改訂部分についても、既存部分と同一の用語を用いることが要求される。

【0024】そこで、本発明においては、このような単語に対して、他の単語よりも高いポイント数を付与して重要単語とし、この重要単語が翻訳文に反映されるようにしている。

【0025】なお、図2に示すテンプレート群では、『dotted』、『there』および『nationwide』が、重要単語として設定されているが、すべての単語のポイント数を「1」にしてもよい。

【0026】ところで、前記翻訳用データベース1には、原文用のテンプレート群とともに、例えば日本語の訳文用テンプレート群も記憶されており、これらは相互に関連付けられている。したがって、図2に示すテンプレート群から得られるテンプレートパターン候補（文章）と、図示しない日本語の訳文用テンプレート群から得られるテンプレートパターン候補（文章）とは、一対一で対応しており、英語のテンプレートパターン候補が特定されれば、これに対応する日本語のテンプレートパターン候補を、日本語の翻訳文として得ることができるようになっている。これは、ATN（拡張遷移ネットワーク）を使用してコンピュータに組み込み格納することにより可能である。

【0027】前記入力文記憶手段2に記憶されている入力文で用いられているすべての単語、およびテンプレートパターン候補記憶手段4に記憶されている各テンプレートパターン候補で用いられているすべての単語は、図

1に示す単語出現頻度記憶手段6において、テンプレート群中における出現頻度が、各テンプレートパターン候補および入力文毎に単語と関連付けて記憶されるようになっており、各テンプレートパターン候補および入力文は、単語の出現頻度とその単語のポイント数に基づき、正規化LSIベクトル演算手段7において、正規化LSI (Latent Semantic Indexing) ベクトルが演算されるようになっている。そして、演算された各テンプレートパターン候補および入力文の正規化LSIベクトルは、正規化LSIベクトル記憶手段8に記憶されるようになっている。

【0028】この正規化LSIベクトル記憶手段8に記憶されている各正規化LSIベクトルは、SVD (特異値ベクトル分解法) アルゴリズムを用いて設定される圧縮文書LSIベクトル空間において相互に比較され、類似度演算手段9において、入力文に対する各テンプレートパターン候補の類似度が演算されるようになっている。そして、類似度が高い方から順に選ばれた複数のテンプレートパターン候補は、優良テンプレートパターン候補として優良テンプレートパターン候補記憶手段10に記憶されるようになっている。

【0029】このようにして得られた各優良テンプレートパターン候補は、図1に示すように、一致単語検索手段11において、入力文と一致する単語が検索されるようになっており、また合計ポイント数演算手段12においては、各優良テンプレートパターン候補で用いられているすべての単語の合計ポイント数が演算されるとともに、一致単語ポイント数演算手段13においては、入力文と各優良テンプレートパターン候補との間での一致単語の合計ポイント数が演算されるようになっている。そして、これら両合計ポイント数は、最適値比較演算手段14において比較され、各優良テンプレートパターン候補の最適値比較値が演算されるようになっている。

【0030】すなわち、最適値比較演算手段14においては、各優良テンプレートパターン候補における全単語の合計ポイント数を分母とするとともに、各優良テンプレートパターン候補における入力文との一致単語の合計ポイント数を分子とすることにより、最適値比較値が演算されるようになっており、各優良テンプレートパターン候補の最適値比較値は、図1に示すように、最適テンプレートパターン候補決定手段15において相互に比較され、最も大きな最適値比較値を有する優良テンプレートパターン候補が、最適テンプレートパターン候補として決定されるようになっている。そして、一致単語の合計ポイント数のみならず、各優良テンプレートパターン候補における全単語の合計ポイント数をも考慮されるので、入力文により類似した優良テンプレートパターン候補が最適テンプレートパターン候補として決定されるようになっている。

【0031】このようにして、決定された最適テンプレ

ートパターン候補は、これに対応する日本語の訳文用テンプレートパターンが、前記翻訳用データベース1の訳文用テンプレート群の中に存在するので、最適翻訳文選定手段16において、この訳文用テンプレートパターン候補が最適翻訳文として選定されるようになっている。

【0032】図3は、前記最適翻訳文選定装置における最適翻訳文選定方法を示す流れ図であり、以下これについて説明する。

【0033】まず、入力文と同一言語、例えば英語の原文用のテンプレート群およびこのテンプレート群と意味的に等価な例えば日本語の訳文用テンプレート群が予め用意され、これら両テンプレート群は、相互に関連付けられた状態となっている。そして、ステップS1においてこれら両テンプレート群を読み込む。これら両テンプレート群は通常、ATTN (拡張遷移ネットワーク) を使用してコンピュータに組み込み格納される。

【0034】次いで、ステップS2において、原文用のテンプレート群中のすべての単語に、その重要度に応じたポイント数で付与されているポイントを読み込む。

【0035】次いで、ステップS3において、入力文の読み込みを行なうとともに、ステップS4において、入力文の解析を行なう。

【0036】なお、この入力文の解析としては、例えば特開平9-325673号公報に示されているように、基本的用法が正しく履行されているか否かを調べるためのキーパターンによるキーパターンチェック、入力文のすべての単語に対し内蔵する辞書を用いて行なわれる品詞分類、および構文解析プログラムによる構文解析が順次行なわれることになる。そしてその後、ステップS5において、テンプレート群の各テンプレートブロックと入力文とを比較し、作成可能なすべてのテンプレートパターン候補を検索する。

【0037】このようにして得られたテンプレートパターン候補の数は、翻訳用データベース1内のデータ量にもよるが、常に精度の高い最適翻訳文が得られる程度に翻訳用データベース1内のデータ量を多くした場合には、しばしばかなりの数となる。

【0038】そこで本発明においては、ステップS6において、テンプレートパターン候補の粗い絞り込みを行なって対象となるテンプレートパターン候補の数を少なくした後、この少ない数のテンプレートパターン候補のみを対象として、ステップS7において、テンプレートパターン候補の精細な絞り込みを行なうようにしている。

【0039】この精細な絞り込みにより、最適テンプレートパターン候補が決定されるので、ステップS8において、この最適テンプレートパターン候補に対応する日本語のテンプレートパターン候補を翻訳用データベース1から選んで、これを最適翻訳文として決定する。

【0040】図4は、図3のステップS6での処理を詳

細に示すもので、以下これについて説明する。まず、ステップS11において、各テンプレートパターン候補および入力文で用いられているすべての単語のテンプレート群中における出現頻度とその単語のポイント数とに基づき、各テンプレートパターン候補および入力文の正規化LSIベクトルを演算し、次いでステップS12において、SVDアルゴリズムを用いて設定される圧縮文書LSIベクトル空間に、前記各正規化LSIベクトルを射影することにより、入力文に対する各テンプレートパターン候補の類似度を演算する。そしてその後、ステップS13において、類似度が高い方から複数のテンプレートパターン候補を選び、これを優良テンプレートパターン候補として決定する。

【0041】なおここで、優良テンプレートパターン候補は、類似度が最も高いものを1つだけ選ぶことも理論的には可能であるが、後に詳述するように、類似度の高さと文書の近似度とは、必ずしも一致しないので、文書の近似度が高いものが必ず優良テンプレートパターン候補として選ばれるようにするためには、複数のテンプレートパターン候補を優良テンプレートパターン候補として決定しておくことが好ましい。

【0042】次に、どのようにして優良テンプレートパターン候補が決定されるのかにつき、例えば米国特許第4839853号に示されている例を用いて具体的に説明する。

【0043】いま、テンプレートパターン候補として、
C1: Human machine interface for LabABC computer applications (研究所ABCコンピュータ・アプリケーション用のヒューマン・マシン・インターフェース)

C2: A survey of user opinion of computer system response time (コンピュータ・システム応答時間に関するユーザの意見調査)

C3: The EPS user interface management system (EPSユーザ・インターフェース管理システム)

C4: Systems and human systems engineering testing of EPS-2 (EPS-2のシステムおよびヒューマン・システム・エンジニアリング・テスト)

C5: Relation of user-perceived response time to error measurement (エラー測定に対するユーザ知覚応答時間の関係)

M1: The generation of random, binary, unordered trees (ランダム木、二進木、無秩序木の生成)

M2: The intersection graph of paths in trees (木の経路を示す交差グラフ)

M3: Graph minors IV: Widths of trees and well-quasi-ordering (グラフ小行列式IV: 木の幅と整列準配列)

M4: Graph minors: A survey (グラフ小行列式: 調査)

の9個のテンプレートパターン候補が検索されたものとする、文書別用語行列Dは表1のようになる。

【0044】

【表1】

用 語	文					書			
	C1	C2	C3	C4	C5	M1	M2	M3	M4
human	1	0	0	1	0	0	0	0	0
interface	1	0	1	0	0	0	0	0	0
computer	1	1	0	0	0	0	0	0	0
user	0	1	1	0	1	0	0	0	0
system	0	1	1	2	0	0	0	0	0
response	0	1	0	0	1	0	0	0	0
time	0	1	0	0	1	0	0	0	0
EPS	0	0	1	1	0	0	0	0	0
survey	0	1	0	0	0	0	0	0	1
tree	0	0	0	0	0	1	1	1	0
graph	0	0	0	0	0	0	1	1	1
minor	0	0	0	0	0	0	0	1	1

【0045】なお、表1において、各用語のポイント数はすべて「1」であり、また表1中の数字は、テンプレート群中における出現頻度としての出現回数である。この出現頻度は、生の出現回数を用いずに、出現回数のブールまたは対数をとってもよい。

【0046】次に、特異値ベクトル分解 $D=USV^T$ に

よって、用語行列 U 、対角行列 S 、文書行列 V を求めると、これら各行列 U 、 S 、 V は、表2、表3および表4のようになる。

【0047】

【表2】

用語行列 (12個の用語×9次元)									
human	0.22	-0.11	0.29	-0.41	-0.11	-0.34	-0.52	-0.06	-0.41
interface	0.20	-0.07	0.14	-0.55	0.28	0.50	-0.07	-0.01	-0.11
computer	0.24	0.04	-0.16	-0.59	-0.11	-0.25	-0.30	0.06	0.49
user	0.40	0.06	-0.34	0.10	0.33	0.38	0.00	0.00	0.01
system	0.64	-0.17	0.36	0.33	-0.16	-0.21	-0.16	0.03	0.27
response	0.26	0.11	-0.42	0.07	0.08	-0.17	0.28	-0.02	-0.05
time	0.26	0.11	-0.42	0.07	0.08	-0.17	0.28	-0.02	-0.05
EPS	0.30	-0.14	0.33	0.19	0.11	0.27	0.03	-0.02	-0.16
survey	0.20	0.27	-0.18	-0.03	-0.54	0.08	-0.47	-0.04	-0.58
tree	0.01	0.49	0.23	0.02	0.59	-0.39	-0.29	0.25	-0.22
graph	0.04	0.62	0.22	0.00	-0.07	0.11	0.16	-0.68	0.23
minor	0.03	0.45	0.14	-0.01	-0.30	0.28	0.34	0.68	0.18

【0048】

【表3】

対角行列 (9個の特異値)									
3.34									
	2.54								
		2.35							
			1.64						
				1.50					
					1.31				
						0.84			
							0.56		
								0.36	

【0049】

【表4】

文書行列 (9個の文書×9次元)									
C 1	0.20	-0.06	0.11	-0.95	0.04	-0.08	0.18	-0.01	-0.06
C 2	0.60	0.18	-0.50	-0.03	-0.21	-0.02	-0.43	0.05	0.24
C 3	0.46	-0.13	0.21	0.04	0.38	0.07	-0.24	0.01	0.02
C 4	0.54	-0.23	0.57	0.27	-0.20	-0.04	0.26	-0.02	-0.08
C 5	0.28	0.11	-0.50	0.15	0.33	0.03	0.67	-0.06	-0.26
M 1	0.00	0.19	0.10	0.02	0.39	-0.30	-0.34	0.45	-0.62
M 2	0.01	0.44	0.19	0.02	0.35	-0.21	-0.15	-0.76	0.02
M 3	0.02	0.62	0.25	0.01	0.15	0.00	0.25	0.45	0.52
M 4	0.08	0.53	0.08	-0.02	-0.60	0.36	0.04	-0.07	-0.45

【0050】ところで、文書別用語行列Dの主要関連構造を捉えるには、通常は二次元で充分である。したがって、用語行列Uおよび文書行列Vの最初の2列のみについて検討して残りの列は無視し、対角行列S内の2つの特異値のみにつき検討して残りの列を無視するものとする。すると、図5に示すような二次元の圧縮文書LSIベクトル空間が得られるので、この圧縮文書LSIベクトル空間に、各テンプレートパターン候補C1, C2, C3, C4, C5, M1, M2, M3, M4を射影する。

【0051】ここで、入力文が『human computer interaction』であるとして、この入力文のベクトルQは、次式で表すことができる。

【0052】

【数1】

$$Q = [101000000000]^T$$

【0053】このベクトルQを図5に示す圧縮LSIベクトル空間に射影すると、ベクトルQに対応する圧縮文書ベクトルV_Qは、次式のようにになる。

【0054】

【数2】

$$V_Q = Q^T U S^{-1} = [0.14 \quad -0.03]$$

【0055】この圧縮文書ベクトルV_Qは、図5においては符号Qで示すようになるが、大雑把に言えば、入力文は、圧縮文書LSIベクトル空間において、その角度

から見てテンプレートパターン候補C1, C2, C3, C4, C5に極めて近い(類似度が高い)ことが判る。

【0056】なお、厳密に言えば、入力文と各テンプレートパターン候補との比較は、次式で得られる角度による。

【0057】

【数3】

$$(V_1, S) \times (V_2, S)^T$$

$$\cos \theta = \frac{(V_1, S) \times (V_2, S)^T}{\| (V_1, S) \|_2 \times \| (V_2, S) \|_2}$$

$$\| (V_1, S) \|_2 \times \| (V_2, S) \|_2$$

【0058】ところで、類似度が高い各テンプレートパターン候補C1, C2, C3, C4, C5の中には、入力文と用語を全く共有しないテンプレートパターン候補C3, C5も含まれている。したがって、例えばテンプレートパターン候補C3の類似度が最も高くなった場合に、このテンプレートパターン候補C3のみを優良テンプレートパターン候補として決定した場合には、正しい翻訳文が得られないことになる。

【0059】そこで本発明においては、類似度が高い各テンプレートパターン候補C1, C2, C3, C4, C5のすべてを、優良テンプレートパターン候補として決定するようにしている。

【0060】図6は、図3のステップS7での処理を詳細に示すもので、以下これについて説明する。まず、ステップS21において、入力文中の各単語が、各優良テンプレートパターン候補のどの単語と一致するかを検索するとともに、ステップS22において、一致した単語

の合計ポイント数を各優良テンプレートパターン候補毎に演算する。

【0061】次いで、ステップS33において、各優良テンプレートパターン候補で用いられているすべての単語の合計ポイント数を各優良テンプレートパターン候補毎に演算するとともに、ステップS24において、各優良テンプレートパターン候補の最適比較値をそれぞれ演算する。

【0062】この最適比較値は、具体的には、ステップS23で求められた合計ポイント数を分母、ステップS22で求められた一致単語ポイント数を分子とすることにより求められる。すなわち、一致単語ポイント数を合計ポイント数で除すことにより求められる。

【0063】次いで、ステップS25において、各優良テンプレートパターン候補の最適比較値を比較し、ステップS22で求められた一致単語ポイント数が最大となる最適比較値のうちで、最適比較値が最も大きくなる優良テンプレートパターン候補を、最適テンプレートパターン候補として決定する。

【0064】例えば入力文が、

『In Japan, are dotted with lovely public gardens nationwide.』

でありこの入力文に対する最適テンプレートパターン候補を、図2に示すテンプレート群を用いて決定すると、

『Japan is dotted with lovely public gardens nationwide.』

が最適テンプレートパターン候補として決定されることになる。

【0065】次に、前記入力文中の

『In Japan, are dotted with』

の部分に対して、図2に示すテンプレート群から選択された優良テンプレートパターン候補の中から、どのようにして最適テンプレートパターン候補が決定されるのかにつき具体的に説明する。

【0066】いま、

『In Japan, are dotted with』

という入力文に対して、図2に示すテンプレート群から、

『Japan is dotted with』

と、

『In Japan, there are』

とが優良テンプレートパターン候補として選択されたものとすると、図7に示すような有向グラフが得られる。

【0067】この有向グラフのノード（状態）が、N1, N2, N3, N4, N6, N7, N8, N9, N5とソート済みであることを前提とした上で、各ノードと

入力文とのマッチングを行なう。

【0068】図8は初期状態を示し、この状態から、アーク(N1, N2)すなわち『Japan』、アーク(N1, N6)すなわち『In』、アーク(N2, N3)すなわち『is』、アーク(N3, N4)すなわち『dotted』、アーク(N4, N5)すなわち『with』、アーク(N6, N7)すなわち『Japan』、アーク(N7, N8)すなわち『.』、アーク(N8, N9)すなわち『there』、およびアーク(N9, N5)すなわち『are』と、入力文中の各単語とのマッチングが順次行なわれる。

【0069】まず、アーク(N1, N2)と入力文とのマッチングについて、図9を参照して説明する。

(1) アーク(N1, N2)と入力文の『In』とを照合すると、両者は等しくないので、セル(N2, “In”)の値およびセル(N1, “In”)の値をチェックし、両者は共に0であることから、セル(N2, “In”)はそのままとし、バックリンクも同様にそのままとする(セル(N1, “In”)のバックリンクは当初設定されていないので、この時点で(N1, “In”)に設定する)。図9の表中における矢印がバックリンクである。

【0070】ここで、バックリンクとは、以下のような理由で設定されるものである。すなわち、このマッチング方法では、図7に示す始端ノードN1から終端ノードN5までのすべてのマッチング作業を完了しなければ、最適テンプレートパターンがどの経路の文字列なのか判明しない。具体的には、図7の終端ノードN5まで作業を完了した時点で、アーク(N4, N5)側が最適テンプレートパターンであると判明した場合、ノードN4は、N4→N3→N2→N1の経路を辿って始端ノードN1に至る文字列であることが判明する必要がある。この経路を示すものがバックリンクである。

【0071】(2) アーク(N1, N2)と入力文の『Japan』とを照合する。すると、両者は等しいので、セル(N2, “Japan”)の値(=0)、セル(N2, “In”)の値(=0)、セル(N1, “Japan”)の値(=0)、およびセル(N1, “In”) + 1の値(=1、ここで+1は図7の“Japan”のポイント数を加えることを意味する。)をチェックし、そのうちの最大値を選択してセル(N2, “Japan”)の値とする。この数字は、前記最適比較値を求める際の分子の数字として用いられるもので、これにより、セル(N2, “Japan”) = 1となり、バックリンクは、最大値として選択したセル(N1, “In”)にリンクされる(ここで、N1はアーク(N1, N2)の始端ノードである)。なお、図9に示す表において、N2の行の各昇目の右下に記している数字「1」は、図7の始端ノードN1からノードN2までに含まれる単語の合計のポイント数を表わし、この数字は、前記

最適値比較値を求める際の分母の数字として用いられる。

【0072】(3) 次に、アーク(N1, N2)と入力文の『』とを照合する。すると、両者は等しくないで、セル(N2, “,”)の値(=0)、セル(N2, “Japan”)の値(=1)、およびセル(N1, “,”)の値(=0)をチェックし、そのうちの最大値を選択する。したがって、セル(N2, “,”)=1となり、バックリンクは、最大値として選択したセル(N2, “Japan”)にリンクされる。

【0073】(4) 次に、アーク(N1, N2)と入力文の『are』とを照合する。すると、両者は等しくないで、セル(N2, “are”)の値(=0)、セル(N2, “,”)の値(=1)、およびセル(N1, “are”)の値(=0)をチェックし、そのうちの最大値を選択する。したがって、セル(N2, “are”)=1となり、バックリンクは、最大値として選択したセル(N2, “,”)にリンクされる。

【0074】(5) 次に、アーク(N1, N2)と入力文の『dotted』とを照合する。すると、両者は等しくないで、セル(N2, “dotted”)の値(=0)、セル(N2, “are”)の値(=1)、およびセル(N1, “dotted”)の値(=0)をチェックし、そのうちの最大値を選択する。したがって、セル(N2, “dotted”)=1となり、バックリンクは、最大値として選択したセル(N2, “are”)にリンクされる。

【0075】(6) 次に、アーク(N1, N2)と入力文の『with』とを照合する。すると、両者は等しくないで、セル(N2, “with”)の値(=0)、セル(N2, “dotted”)の値(=1)、およびセル(N1, “with”)の値(=0)をチェックし、そのうちの最大値を選択する。したがって、セル(N2, “with”)=1となり、バックリンクは、最大値として選択したセル(N2, “dotted”)にリンクされる。

【0076】次に、アーク(N1, N6)と入力文とのマッチングについて、図10を参照して説明する。

(1) まず、アーク(N1, N6)と入力文の『In』とを照合すると、両者は等しいので、セル(N6, “In”)の値(=0)、およびセル(N1, “In”)+1の値(=1、ここで+1は図7の“In”のポイント数を加えることを意味する。)をチェックし、そのうちの最大値を選択してセル(N6, “In”)の値とする。これにより、セル(N6, “In”)=1となり、バックリンクは、最大値として選択したセル(N1, “In”)にリンクされる。なお、図10に示す表において、N6の行の各列目の右下に記している数字「1」は、図7の始端ノードN1からノードN6までに含まれる単語の合計のポイント数、すなわち『In』の

ポイント数である。

【0077】(2) 次に、アーク(N1, N6)と入力文の『Japan』とを照合する。すると、両者は等しくないで、セル(N6, “Japan”)の値(=0)、セル(N6, “In”)の値(=1)、およびセル(N1, “Japan”)の値(=0)をチェックし、そのうちの最大値を選択する。したがって、セル(N6, “Japan”)=1となり、バックリンクは、最大値として選択したセル(N6, “In”)にリンクされる。

【0078】(3) 次に、アーク(N1, N6)と入力文の『』とを照合する。すると、両者は等しくないで、セル(N6, “,”)の値(=0)、セル(N6, “Japan”)の値(=1)、およびセル(N1, “,”)の値(=0)をチェックし、そのうちの最大値を選択する。したがって、セル(N6, “,”)=1となり、バックリンクは、最大値として選択したセル(N6, “Japan”)にリンクされる。

【0079】(4) 次に、アーク(N1, N6)と入力文の『are』とを照合する。すると、両者は等しくないで、セル(N6, “are”)の値(=0)、セル(N6, “,”)の値(=1)、およびセル(N1, “are”)の値(=0)をチェックし、そのうちの最大値を選択する。したがって、セル(N6, “are”)=1となり、バックリンクは、最大値として選択したセル(N6, “,”)にリンクされる。

【0080】(5) 次に、アーク(N1, N6)と入力文の『dotted』とを照合する。すると、両者は等しくないで、セル(N6, “dotted”)の値(=0)、セル(N6, “are”)の値(=1)、およびセル(N1, “dotted”)の値(=0)をチェックし、そのうちの最大値を選択する。したがって、セル(N6, “dotted”)=1となり、バックリンクは、最大値として選択したセル(N6, “are”)にリンクされる。

【0081】(6) 次に、アーク(N1, N6)と入力文の『with』とを照合する。すると、両者は等しくないで、セル(N6, “with”)の値(=0)、セル(N6, “dotted”)の値(=1)、およびセル(N1, “with”)の値(=0)をチェックし、そのうちの最大値を選択する。したがって、セル(N6, “with”)=1となり、バックリンクは、最大値として選択したセル(N6, “dotted”)にリンクされる。

【0082】次に、アーク(N2, N3)と入力文とのマッチングについて、図11を参照して説明する。

(1) まず、アーク(N2, N3)と入力文の『In』とを照合すると、両者は等しくないで、セル(N3, “In”)の値およびセル(N2, “In”)をチェックし、両者は共に0であるので、セル(N3, “I

n”）はそのままとし、バックリンクもそのままとする。なお、図11に示す表において、N3の行の各升目の右下に記している数字「2」は、図7の始端ノードN1からノードN3までに含まれる単語の合計のポイント数、すなわち『Japan』および『is』の合計のポイント数である。

【0083】(2) 次に、アーク(N2, N3)と入力文の『Japan』とを照合する。すると、両者は等しくないで、セル(N3, “Japan”)の値(=0)、セル(N3, “In”)の値(=0)、およびセル(N2, “Japan”)の値(=1)をチェックし、そのうちの最大値を選択する。したがって、セル(N3, “Japan”) = 1となり、バックリンクは、最大値として選択したセル(N2, “Japan”)にリンクされる。

【0084】(3) 次に、アーク(N2, N3)と入力文の『.』とを照合する。すると、両者は等しくないで、セル(N3, “.”)の値(=0)、セル(N3, “Japan”)の値(=1)、およびセル(N2, “.”)の値(=1)をチェックし、そのうちの最大値を選択する。したがって、セル(N3, “.”) = 1となり、バックリンクは、最大値として選択したセル(N3, “Japan”)にリンクされる。

【0085】(4) 次に、アーク(N2, N3)と入力文の『are』とを照合する。すると、両者は等しくないで、セル(N3, “are”)の値(=0)、セル(N3, “.”)の値(=1)、およびセル(N2, “are”)の値(=1)をチェックし、そのうちの最大値を選択する。したがって、セル(N3, “are”) = 1となり、バックリンクは、最大値として選択したセル(N3, “.”)にリンクされる。

【0086】(5) 次に、アーク(N2, N3)と入力文の『dotted』とを照合する。すると、両者は等しくないで、セル(N3, “dotted”)の値(=0)、セル(N3, “are”)の値(=1)、およびセル(N2, “dotted”)の値(=1)をチェックし、そのうちの最大値を選択する。したがって、セル(N3, “dotted”) = 1となり、バックリンクは、最大値として選択したセル(N3, “are”)にリンクされる。

【0087】(6) 次に、アーク(N2, N3)と入力文の『with』とを照合する。すると、両者は等しくないで、セル(N3, “with”)の値(=0)、セル(N3, “dotted”)の値(=1)、およびセル(N2, “with”)の値(=1)をチェックし、そのうちの最大値を選択する。したがって、セル(N3, “with”) = 1となり、バックリンクは、最大値として選択したセル(N3, “dotted”)にリンクされる。

【0088】次に、アーク(N3, N4)と入力文との

マッチングについて、図12を参照して説明する。

(1) まず、アーク(N3, N4)と入力文の『In』とを照合すると、両者は等しくないで、セル(N4, “In”)の値およびセル(N3, “In”)の値をチェックし、両者は共に0であるので、セル(N4, “In”)はそのままとし、バックリンクもそのままとする。なお、図12に示す表において、N4の行の各升目の右下に記している数字「4」は、図7の始端ノードN1からノードN4までに含まれる単語、すなわち『Japan』, 『is』および『dotted』の合計のポイント数である。

【0089】(2) 次に、アーク(N3, N4)と入力文の『Japan』とを照合する。すると、両者は等しくないで、セル(N4, “Japan”)の値(=0)、セル(N4, “In”)の値(=0)、およびセル(N3, “Japan”)の値(=1)をチェックし、そのうちの最大値を選択する。したがって、セル(N4, “Japan”) = 1となり、バックリンクは、最大値として選択したセル(N3, “Japan”)にリンクされる。

【0090】(3) 次に、アーク(N3, N4)と入力文の『.』とを照合する。すると、両者は等しくないで、セル(N4, “.”)の値(=0)、セル(N4, “Japan”)の値(=1)、およびセル(N3, “.”)の値(=1)をチェックし、そのうちの最大値を選択する。したがって、セル(N4, “.”) = 1となり、バックリンクは、最大値として選択したセル(N4, “Japan”)にリンクされる。

【0091】(4) 次に、アーク(N3, N4)と入力文の『are』とを照合する。すると、両者は等しくないで、セル(N4, “are”)の値(=0)、セル(N4, “.”)の値(=1)、およびセル(N3, “are”)の値(=1)をチェックし、そのうちの最大値を選択する。したがって、セル(N4, “are”) = 1となり、バックリンクは、最大値として選択したセル(N4, “.”)にリンクされる。

【0092】(5) 次に、アーク(N3, N4)と入力文の『dotted』とを照合する。すると、両者は等しいので、セル(N4, “dotted”)の値(=0)、セル(N4, “are”)の値(=1)、セル(N3, “dotted”)の値(=1)、およびセル(N3, “are”) + 2の値(=3、ここで+2は図7の“dotted”のポイント数を加えることを意味する。)をチェックし、その最大値を選択する。したがって、セル(N4, “dotted”) = 3となり、バックリンクは、最大値として選択したセル(N3, “are”)にリンクされる。

【0093】(6) 次に、アーク(N3, N4)と入力文の『with』とを照合する。すると、両者は等しくないで、セル(N4, “with”)の値(=

0)、セル(N4, "dotted")の値(=3)、およびセル(N3, "with")の値(=1)をチェックし、そのうちの最大値を選択する。したがって、セル(N4, "with")=3となり、バックリンクは、最大値として選択したセル(N4, "dotted")にリンクされる。

【0094】次に、アーク(N4, N5)と入力文とのマッチングについて、図13を参照して説明する。

(1) まず、アーク(N4, N5)と入力文の『In』とを照合すると、両者は等しくないので、セル(N5, "In")の値およびセル(N4, "In")の値をチェックし、両者は共に0であるので、セル(N5, "In")はそのままとし、バックリンクもそのままとする。なお、図13に示す表において、N5の行の各升目の右下に記している数字「5」は、図7の始端ノードN1からノードN5までの上側の文字列に含まれる単語、すなわち『Japan』, 『is』, 『dotted』および『with』の合計のポイント数である。

【0095】(2) 次に、アーク(N4, N5)と入力文の『Japan』とを照合する。すると、両者は等しくないで、セル(N5, "Japan")の値(=0)、セル(N5, "In")の値(=0)、およびセル(N4, "Japan")の値(=1)をチェックし、そのうちの最大値を選択する。したがって、セル(N5, "Japan")=1となり、バックリンクは、最大値として選択したセル(N4, "Japan")にリンクされる。

【0096】(3) 次に、アーク(N4, N5)と入力文の『,』とを照合する。すると、両者は等しくないで、セル(N5, ",")の値(=0)、セル(N5, "Japan")の値(=1)、およびセル(N4, ",")の値(=1)をチェックし、そのうちの最大値を選択する。したがって、セル(N5, ",")=1となり、バックリンクは、最大値として選択したセル(N5, "Japan")にリンクされる。

【0097】(4) 次に、アーク(N4, N5)と入力文の『are』とを照合する。すると、両者は等しくないで、セル(N5, "are")の値(=0)、セル(N5, ",")の値(=1)、およびセル(N4, "are")の値(=1)をチェックし、そのうちの最大値を選択する。したがって、セル(N5, "are")=1となり、バックリンクは、最大値として選択したセル(N5, ",")にリンクされる。

【0098】(5) 次に、アーク(N4, N5)と入力文の『dotted』とを照合する。すると、両者は等しくないで、セル(N5, "dotted")の値(=0)、セル(N5, "are")の値(=1)、およびセル(N4, "dotted")の値(=3)をチェックし、そのうちの最大値を選択する。したがって、セル(N5, "dotted")=3となり、バックリ

nkは、最大値として選択したセル(N4, "dotted")にリンクされる。

【0099】(6) 次に、アーク(N4, N5)と入力文の『with』とを照合する。すると、両者は等しいので、セル(N5, "with")の値(=0)、セル(N5, "dotted")の値(=3)、セル(N4, "with")の値(=3)、およびセル(N4, "dotted")+1の値(=4、ここで+1は図7の"with"のポイント数を加えることを意味する。)をチェックし、そのうちの最大値を選択する。したがって、セル(N5, "with")=4となり、バックリンクは、最大値として選択したセル(N4, "dotted")にリンクされる。

【0100】次に、アーク(N6, N7)と入力文とのマッチングについて、図14を参照して説明する。

(1) まず、アーク(N6, N7)と入力文の『In』とを照合すると、両者は等しくないで、セル(N7, "In")の値(=0)とセル(N6, "In")の値(=1)とをチェックし、そのうちの大きい方の値を選択する。したがって、セル(N7, "In")=1となり、バックリンクは、選択したセル(N6, "In")にリンクされる。なお、図14に示す表において、N7の行の各升目の右下に記している数字「2」は、図7の始端ノードN1からノードN7までに含まれる単語、すなわち『In』および『Japan』の合計のポイント数である。

【0101】(2) 次に、アーク(N6, N7)と入力文の『Japan』とを照合する。すると、両者は等しいので、セル(N7, "Japan")の値(=0)、セル(N7, "In")の値(=0)、セル(N6, "Japan")の値(=1)、およびセル(N6, "In")+1の値(=2、ここで+1は図7の"Japan"のポイント数を加えることを意味する。)をチェックし、そのうちの最大値を選択する。したがって、セル(N7, "Japan")=2となり、バックリンクは、最大値として選択したセル(N6, "In")にリンクされる。

【0102】(3) 次に、アーク(N6, N7)と入力文の『,』とを照合する。すると、両者は等しくないで、セル(N7, ",")の値(=0)、セル(N7, "Japan")の値(=2)、およびセル(N6, ",")の値(=1)をチェックし、そのうちの最大値を選択する。したがって、セル(N7, ",")=2となり、バックリンクは、最大値として選択したセル(N7, "Japan")にリンクされる。

【0103】(4) 次に、アーク(N6, N7)と入力文の『are』とを照合する。すると、両者は等しくないで、セル(N7, "are")の値(=0)、セル(N7, ",")の値(=2)、およびセル(N6, "are")の値(=1)をチェックし、そのうちの最

大値を選択する。したがって、セル(N7, “are”) = 2となり、バックリンクは、最大値として選択したセル(N7, “.”)にリンクされる。

【0104】(5) 次に、アーク(N6, N7)と入力文の『dotted』とを照合する。すると、両者は等しくないので、セル(N7, “dotted”)の値(=0)、セル(N7, “are”)の値(=2)、およびセル(N6, “dotted”)の値(=1)をチェックし、そのうちの最大値を選択する。したがって、セル(N7, “dotted”) = 2となり、バックリンクは、最大値として選択したセル(N7, “are”)にリンクされる。

【0105】(6) 次に、アーク(N6, N7)と入力文の『with』とを照合する。すると、両者は等しくないので、セル(N7, “with”)の値(=0)、セル(N7, “dotted”)の値(=2)、およびセル(N6, “with”)の値(=1)をチェックし、そのうちの最大値を選択する。したがって、セル(N7, “with”) = 2となり、バックリンクは、最大値として選択したセル(N7, “dotted”)にリンクされる。

【0106】次に、アーク(N7, N8)と入力文とのマッチングについて、図15を参照して説明する。

(1) まず、アーク(N7, N8)と入力文の『In』とを照合すると、両者は等しくないので、セル(N8, “In”)の値(=0)とセル(N7, “In”)の値(=1)とをチェックし、そのうちの大きい方の値を選択する。したがって、セル(N8, “In”) = 1となり、バックリンクは、選択したセル(N7, “In”)にリンクされる。なお、図15に示す表において、N8の行の各昇目の右下に記している数字「3」は、図7の始端ノードN1からノードN8までに含まれる単語、すなわち『In』, 『Japan』および『.』の合計のポイント数である。

【0107】(2) 次に、アーク(N7, N8)と入力文の『Japan』とを照合する。すると、両者は等しくないので、セル(N8, “Japan”)の値(=0)、セル(N8, “In”)の値(=1)、およびセル(N7, “Japan”)の値(=2)をチェックし、そのうちの最大値を選択する。したがって、セル(N8, “Japan”) = 2となり、バックリンクは、最大値として選択したセル(N7, “Japan”)にリンクされる。

【0108】(3) 次に、アーク(N7, N8)と入力文の『.』とを照合する。すると、両者は等しいので、セル(N8, “.”)の値(=0)、セル(N8, “Japan”)の値(=2)、セル(N7, “.”)の値(=2)、およびセル(N7, “Japan”) + 1の値(=3、ここで+1は図7の“.”のポイント数を加えることを意味する。)をチェックし、そのうちの

最大値を選択する。したがって、セル(N8, “.”) = 3となり、バックリンクは、最大値として選択したセル(N7, “Japan”)にリンクされる。

【0109】(4) 次に、アーク(N7, N8)と入力文の『are』とを照合する。すると、両者は等しくないので、セル(N8, “are”)の値(=0)、セル(N8, “.”)の値(=3)、およびセル(N7, “are”)の値(=2)をチェックし、そのうちの最大値を選択する。したがって、セル(N8, “are”) = 3となり、バックリンクは、最大値として選択したセル(N8, “.”)にリンクされる。

【0110】(5) 次に、アーク(N7, N8)と入力文の『dotted』とを照合する。すると、両者は等しくないので、セル(N8, “dotted”)の値(=0)、セル(N8, “are”)の値(=3)、およびセル(N7, “dotted”)の値(=2)をチェックし、そのうちの最大値を選択する。したがって、セル(N8, “dotted”) = 3となり、バックリンクは、最大値として選択したセル(N8, “are”)にリンクされる。

【0111】(6) 次に、アーク(N7, N8)と入力文の『with』とを照合する。すると、両者は等しくないので、セル(N8, “with”)の値(=0)、セル(N8, “dotted”)の値(=3)、およびセル(N7, “with”)の値(=2)をチェックし、そのうちの最大値を選択する。したがって、セル(N8, “with”) = 3となり、バックリンクは、最大値として選択したセル(N8, “dotted”)にリンクされる。

【0112】次に、アーク(N8, N9)と入力文とのマッチングについて、図16を参照して説明する。

(1) まず、アーク(N8, N9)と入力文の『In』とを照合すると、両者は等しくないので、セル(N9, “In”)の値(=0)とセル(N8, “In”)の値(=1)とをチェックし、そのうちの大きい方の値を選択する。したがって、セル(N9, “In”) = 1となり、バックリンクは、選択したセル(N8, “In”)にリンクされる。なお、図16に示す表において、N9の行の各昇目の右下に記している数字「5」は、図7の始端ノードN1からノードN9までに含まれる単語、すなわち『In』, 『Japan』, 『.』および『there』の合計ポイント数である。

【0113】(2) 次に、アーク(N8, N9)と入力文の『Japan』とを照合する。すると、両者は等しくないので、セル(N9, “Japan”)の値(=0)、セル(N9, “In”)の値(=1)、およびセル(N8, “Japan”)の値(=2)をチェックし、そのうちの最大値を選択する。したがって、セル(N9, “Japan”) = 2となり、バックリンクは、最大値として選択したセル(N8, “Japan”)にリンクされる。

n”）にリンクされる。

【0114】（3） 次に、アーク（N8，N9）と入力文の『.』とを照合する。すると、両者は等しくないで、セル（N9，“.”）の値（＝0）、セル（N9，“Japan”）の値（＝2）、およびセル（N8，“.”）の値（＝3）をチェックし、そのうちの最大値を選択する。したがって、セル（N9，“.”）＝3となり、バックリンクは、最大値として選択したセル（N8，“.”）にリンクされる。

【0115】（4） 次に、アーク（N8，N9）と入力文の『are』とを照合する。すると、両者は等しくないで、セル（N9，“are”）の値（＝0）、セル（N9，“.”）の値（＝3）、およびセル（N8，“are”）の値（＝3）をチェックし、そのうちの最大値を選択する。したがって、セル（N9，“are”）＝3となり、バックリンクは、最大値として選択したセル（N9，“.”）にリンクされる。

【0116】（5） 次に、アーク（N8，N9）と入力文の『dotted』とを照合する。すると、両者は等しくないで、セル（N9，“dotted”）の値（＝0）、セル（N9，“are”）の値（＝3）、およびセル（N8，“dotted”）の値（＝3）をチェックし、そのうちの最大値を選択する。したがって、セル（N9，“dotted”）＝3となり、バックリンクは、最大値として選択したセル（N9，“are”）にリンクされる。

【0117】（6） 次に、アーク（N8，N9）と入力文の『with』とを照合する。すると、両者は等しくないで、セル（N9，“with”）の値（＝0）、セル（N9，“dotted”）の値（＝3）、およびセル（N8，“with”）の値（＝3）をチェックし、そのうちの最大値を選択する。したがって、セル（N9，“with”）＝3となり、バックリンクは、最大値として選択したセル（N9，“dotted”）にリンクされる。

【0118】次に、アーク（N9，N5）と入力文とのマッチングについて、図17を参照して説明する。

（1） まず、アーク（N9，N5）と入力文の『In』とを照合すると、両者は等しくないで、セル（N9，“In”）の値（＝0）とセル（N9，“In”）の値（＝1）とをチェックし、そのうちの大きい方の値を選択する。したがって、セル（N5，“In”）＝1となり、バックリンクは、選択したセル（N9，“In”）にリンクされ、バックリンクが変更される。なお、図17に示す表において、N5の行の第1列目から第5列目の各升目の右下に記している数字「6」は、図7の始端ノードN1からノードN5までの下側の文字列に含まれる単語、すなわち『In』、『Japan』、『there』および『are』の合計ポイント数である。ここで、N5の行の第6列目の升目の右

下に記している数字は、「6」ではなく「5」のままであるが、その理由については後に詳述する。

【0119】（2） 次に、アーク（N9，N5）と入力文の『Japan』とを照合する。すると、両者は等しくないで、セル（N5，“Japan”）の値（＝1）、セル（N5，“In”）の値（＝1）、およびセル（N9，“Japan”）の値（＝2）をチェックし、そのうちの最大値を選択する。したがって、セル（N5，“Japan”）＝2となり、バックリンクは、最大値として選択したセル（N9，“Japan”）にリンクされ、バックリンクが変更される。

【0120】（3） 次に、アーク（N9，N5）と入力文の『.』とを照合する。すると、両者は等しくないで、セル（N5，“In”）の値（＝1）、セル（N5，“Japan”）の値（＝2）、およびセル（N9，“.”）の値（＝3）をチェックし、そのうちの最大値を選択する。したがって、セル（N5，“.”）＝3となり、バックリンクは、最大値として選択したセル（N9，“.”）にリンクされ、バックリンクが変更される。

【0121】（4） 次に、アーク（N9，N5）と入力文の『are』とを照合する。すると、両者は等しいので、セル（N5，“are”）の値（＝1）、セル（N5，“.”）の値（＝3）、セル（N9，“are”）の値（＝3）、およびセル（N9，“.”）＋1の値（＝4、ここで＋1は図7の“are”のポイント数を加えることを意味する。）をチェックし、そのうちの最大値を選択する。したがって、セル（N5，“are”）＝4となり、バックリンクは、最大値として選択したセル（N9，“.”）にリンクされ、バックリンクが変更される。

【0122】（5） 次に、アーク（N9，N5）と入力文の『dotted』とを照合する。すると、両者は等しくないで、セル（N5，“dotted”）の値（＝3）、セル（N5，“are”）の値（＝4）、およびセル（N9，“dotted”）の値（＝3）をチェックし、そのうちの最大値を選択する。したがって、セル（N5，“dotted”）＝4となり、バックリンクは、最大値として選択したセル（N5，“are”）にリンクされ、バックリンクが変更される。

【0123】（6） 次に、アーク（N9，N5）と入力文の『with』とを照合する。すると、両者は等しくないで、セル（N5，“with”）の値（＝4）、セル（N5，“dotted”）の値（＝4）、およびセル（N9，“with”）の値（＝3）をチェックする。

【0124】ここで、セル（N5，“dotted”）の値（＝4）が最大値である場合には、N5の行の第1列から第5列までの場合と同様、バックリンクが変更され、各升目の右下に記している数字も、変更された文字

列の合計ポイント数である「6」に変更されることになる。

【0125】ところが、セル(N5, “dotted”)の値(=4)と、元のセル(N5, “with”)の値(=4)とは同一値である。そしてこの場合には、バックリンクを変更することにより、N5の行の升目の右下に記している数字が小さくなる場合を除き、バックリンクの変更は行なわない。したがって、セル(N5, “with”) = 4となり、バックリンクは、セル(N4, “dotted”)にリンクされたままの状態が保持され、升目の右下に記している数字も、図7の上側の文字列の合計ポイント数である「5」のままの状態が保持される。

【0126】以上説明したマッチング作業により、図7に示す上側の文字列の最適比較値が4/5で、下側の文字列の最適比較値が4/6となり、上側の文字列が入力文に対する最適テンプレートパターン候補として決定されることになる。

【0127】このようにして、入力文『In Japan, are dotted with lovely public gardens nationwide.』に対し、『Japan is dotted with lovely public gardens nationwide.』が最適テンプレートパターン候補として決定されたならば、図3のステップS8において、最適テンプレートパターン候補に対応する日本語の訳文用テンプレートパターン候補『日本には、全国的に美しい庭が多い。』が、図1の翻訳用データベース1から選択され、これが最適訳文として決定される。

【0128】しかして、正規化LSIベクトルを圧縮文書LSIベクトル空間に射影する方法を用い、多数のテンプレートパターン候補に対し粗い選定を行なってテンプレートパターン候補の対象数を絞り、次いで絞られた少数のテンプレートパターン候補に対し、最適比較値を用いた精細選定を行なうようにしているので、精度の高い最適訳文を短時間で得ることができる。

【0129】なお、翻訳用データベース1に、十分な量のデータが記憶されている場合には、常に精度の高い最適訳文を得ることができるが、翻訳用データベース1内のデータ量が充分でなかったり、記憶されているデータが適切でない場合には、精度の高い最適訳文が得られないこともある。

【0130】そこで、このような場合には、言語専門家の力およびV T A T (ビジュアル・テンプレート・オーサリング・ツール)と呼ばれているテンプレート・エディタを使用し、翻訳用データベース1の修正とメンテナンスとを行なう。僅かな相違は、翻訳メモリのローカリゼーション技法の場合のように、置換操作によって処理することができる。

【0131】

【発明の効果】以上説明したように本発明は、A T N機能を有する任意数の文章要素を同列で包含する複数のテンプレートブロックで構成される原文用のテンプレート群およびこのテンプレート群と意味的に等価な訳文用テンプレート群が、相互に関連付けて記憶されている翻訳用データベースを用い、前記テンプレート群から作成可能なすべてのテンプレートパターン候補の中から、入力文に最も適したテンプレートパターン候補を検索するとともに、このテンプレートパターン候補に対応する訳文用テンプレートパターン候補を、前記訳文用テンプレート群から作成し、この訳文用テンプレートパターン候補を、最適訳文として選定する最適訳文選定方法であって、前記テンプレート群中のすべての単語に、その重要度に応じたポイント数のポイントを付与した後、前記すべてのテンプレートパターン候補および入力文について、各テンプレートパターン候補および入力文で用いられているすべての単語のテンプレート群中における出現頻度とその単語のポイント数とに基づき、各テンプレートパターン候補および入力文の正規化LSIベクトルを求めるとともに、SVDアルゴリズムを用いて設定される圧縮文書LSIベクトル空間に射影し、入力文に対し圧縮文書LSIベクトル空間における類似度が高い順に複数のテンプレートパターン候補を選んで優良テンプレートパターン候補とし、次いで各優良テンプレートパターン候補で用いられているすべての単語の合計ポイント数を分母、入力文と各優良テンプレートパターン候補との間での一致単語の合計ポイント数を分子とする最適比較値を求め、分子が最大となる最適比較値のうちで、最適比較値が最も大きくなる優良テンプレートパターン候補を選ぶとともに、これに対応する訳文用テンプレートパターン候補を、最適訳文として選定するようにしたことを特徴とする。そして本発明においては、正規化LSIベクトルを圧縮文書LSIベクトル空間に射影する方法を用い、粗い選定を行なって対象となるテンプレートパターンの数を絞り、その後絞られた少数のテンプレートパターンに対し、最適比較値を用いた詳細選定を行なうようにしているので、精度の高い最適訳文を短時間で得ることができる。

【0132】本発明はまた、A T N機能を有する任意数の文章要素を同列で包含する複数のテンプレートブロックで構成される原文用のテンプレート群およびこのテンプレート群と意味的に等価な訳文用テンプレート群とを、相互に関連付けて記憶する翻訳用データベースと；入力文を記憶する入力文記憶手段と；前記テンプレート群の各テンプレートブロックと入力文とを比較し、テンプレート群から作成可能なすべてのテンプレートパターン候補を検索するテンプレートパターン候補検索手段；検索されたテンプレートパターン候補を記憶するテンプレートパターン候補記憶手段と；テンプレート群中のすべての単語にその重要度に応じたポイント数で付与され

るポイントを、単語と関連付けて記憶する単語ポイント数記憶手段と；各テンプレートパターン候補および入力文で用いられるすべての単語のテンプレート群における出現頻度を、各テンプレートパターン候補および入力文毎に単語と関連付けて記憶する単語出現頻度記憶手段と；単語の出現頻度とポイント数とに基づき、各テンプレートパターン候補および入力文の正規化 LSI ベクトルを演算する正規化 LSI ベクトル演算手段と；演算された各テンプレートパターン候補および入力文の正規化 LSI ベクトルを記憶する正規化 LSI ベクトル記憶手段と；SVD アルゴリズムを用いて設定される圧縮文書 LSI ベクトル空間において、各テンプレートパターン候補の正規化 LSI ベクトルと入力文の正規化 LSI ベクトルとを比較し、その類似度を演算する類似度演算手段と；類似度が高い方から順に選ばれた複数のテンプレートパターン候補を、優良テンプレートパターン候補として記憶する優良テンプレートパターン候補記憶手段と；各優良テンプレートパターン候補において入力文と一致する単語を検索する一致単語検索手段と；各優良テンプレートパターン候補で用いられているすべての単語の合計ポイント数を演算する合計ポイント数演算手段と；入力文と各優良テンプレートパターン候補との間での一致単語の合計ポイント数を演算する一致単語ポイント数演算手段と；各優良テンプレートパターン候補における全単語の合計ポイント数を分母、各優良テンプレートパターン候補における入力文との一致単語の合計ポイント数を分子とする最適比較値を演算する最適比較値演算手段と；各優良テンプレートパターン候補の最適比較値を比較し、分子が最大となる最適比較値のうちで、最も大きな最適比較値を有する優良テンプレートパターン候補を、最適テンプレートパターン候補として決定する最適テンプレートパターン候補決定手段と；最適テンプレートパターン候補に対応する訳文用テンプレートパターン候補を、最適翻訳文として選定する最適翻訳文選定手段と；を設け、類似度演算手段等を用い、対象となるテンプレートパターンの数を優良テンプレートパターン候補として絞った後、最適テンプレートパターン候補決定手段等を用い、最適なテンプレートパターン候補を決定するようにしているので、短時間で精度の高い最適翻訳文を選定することができる。

【0133】本発明はさらに、ATN 機能を有する任意数の文章要素を同列で包含する複数のテンプレートブロックで構成される原文用のテンプレート群およびこのテンプレート群と意味的に等価な訳文用テンプレート群を、相互に関連付けてメモリ領域に記憶させる処理；入力文をメモリ領域に記憶させる処理；前記テンプレート群中のすべての単語に対し、その重要度に応じたポイント数で付与されたポイントを、単語と関連付けてメモリ領域に記憶させる処理；前記テンプレート群から作成可能なすべてのテンプレートパターン候補および入力文に

ついて、各テンプレートパターン候補および入力文で用いられているすべての単語のテンプレート群における出現頻度とその単語のポイント数とに基づき、各テンプレートパターン候補および入力文の正規化 LSI ベクトルを求める処理；SVD アルゴリズムを用いて設定される圧縮文書 LSI ベクトル空間において、各テンプレートパターン候補の正規化 LSI ベクトルと入力文の正規化 LSI ベクトルとを比較し、その類似度を求める処理；類似度の高い方から順に選んだ複数のテンプレートパターン候補を、優良テンプレートパターン候補としてメモリ領域に記憶させる処理；各優良テンプレートパターン候補で用いられるすべての単語の合計ポイント数を分母、入力文と各優良テンプレートパターン候補との間での一致単語の合計ポイント数を分子とする最適比較値を用い、分子が最大となる最適比較値のうちで、最適比較値が最も大きくなる優良テンプレートパターン候補を、最適テンプレートパターン候補として決定する処理；最適テンプレートパターン候補に対応する訳文用テンプレートパターン候補を、最適翻訳文として選定する処理；をコンピュータに実行させるようにしているので、前記各処理により、精度の高い最適翻訳文を短時間で得ることができる。

【図面の簡単な説明】

【図 1】本発明の実施の一形態に係る最適翻訳文選定装置を示す全体構成図である。

【図 2】図 1 の装置で用いられるテンプレート群の一例を示す説明図である。

【図 3】図 1 の装置による最適翻訳文選定方法を示す流れ図である。

【図 4】図 3 のステップ S 6 の詳細を示す流れ図である。

【図 5】圧縮文書 LSI ベクトル空間に射影されたテンプレートパターン候補および入力文の状態を示す説明図である。

【図 6】図 3 のステップ S 7 の詳細を示す流れ図である。

【図 7】有向グラフの一例を示す説明図である。

【図 8】図 4 の各ノードと入力文とのマッチングを行なう際の初期状態を示す表である。

【図 9】アーク (N1, N2) と入力文とのマッチングを示す表である。

【図 10】アーク (N1, N6) と入力文とのマッチングを示す表である。

【図 11】アーク (N2, N3) と入力文とのマッチングを示す表である。

【図 12】アーク (N3, N4) と入力文とのマッチングを示す表である。

【図 13】アーク (N4, N5) と入力文とのマッチングを示す表である。

【図 14】アーク (N6, N7) と入力文とのマッチン

グを示す表である。

【図15】アーク(N7, N8)と入力文とのマッチングを示す表である。

【図16】アーク(N8, N9)と入力文とのマッチングを示す表である。

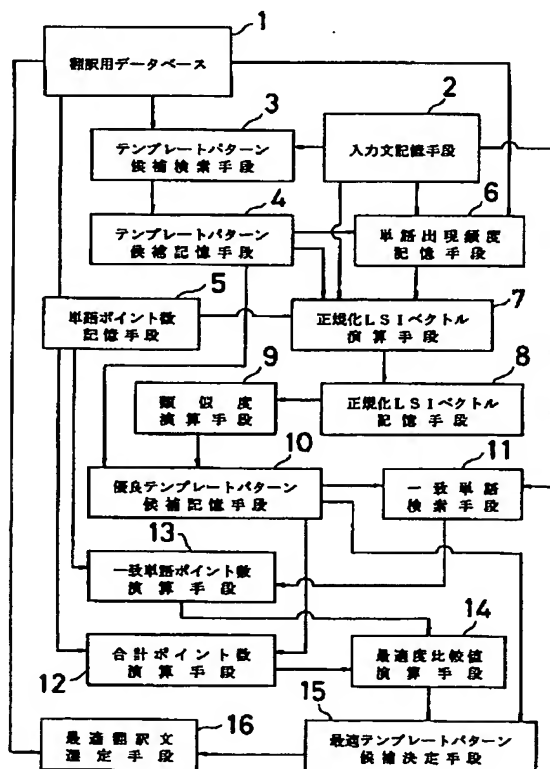
【図17】アーク(N9, N5)と入力文とのマッチングを示す表である。

【符号の説明】

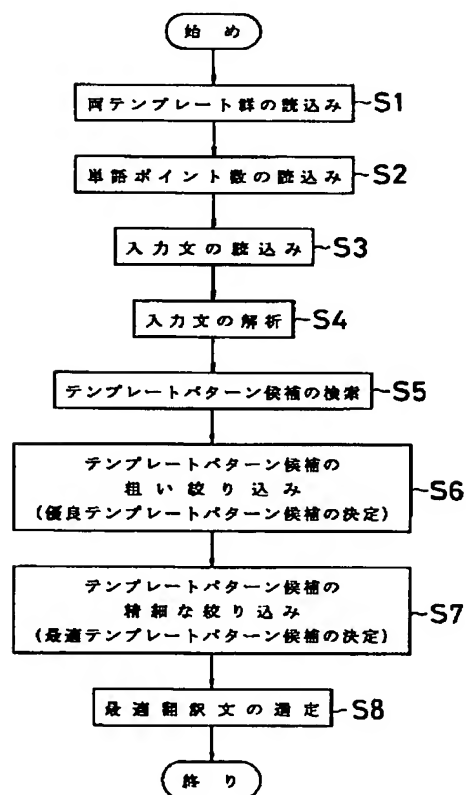
- 1 翻訳用データベース
- 2 入力文記憶手段
- 3 テンプレートパターン候補検索手段
- 4 テンプレートパターン候補記憶手段

- 5 単語ポイント数記憶手段
- 6 単語出現頻度記憶手段
- 7 正規化LSIベクトル演算手段
- 8 正規化LSIベクトル記憶手段
- 9 類似度演算手段
- 10 優良テンプレートパターン候補記憶手段
- 11 一致単語検索手段
- 12 合計ポイント数演算手段
- 13 一致単語ポイント数演算手段
- 14 最適値比較値演算手段
- 15 最適テンプレートパターン候補決定手段
- 16 最適翻訳文選定手段

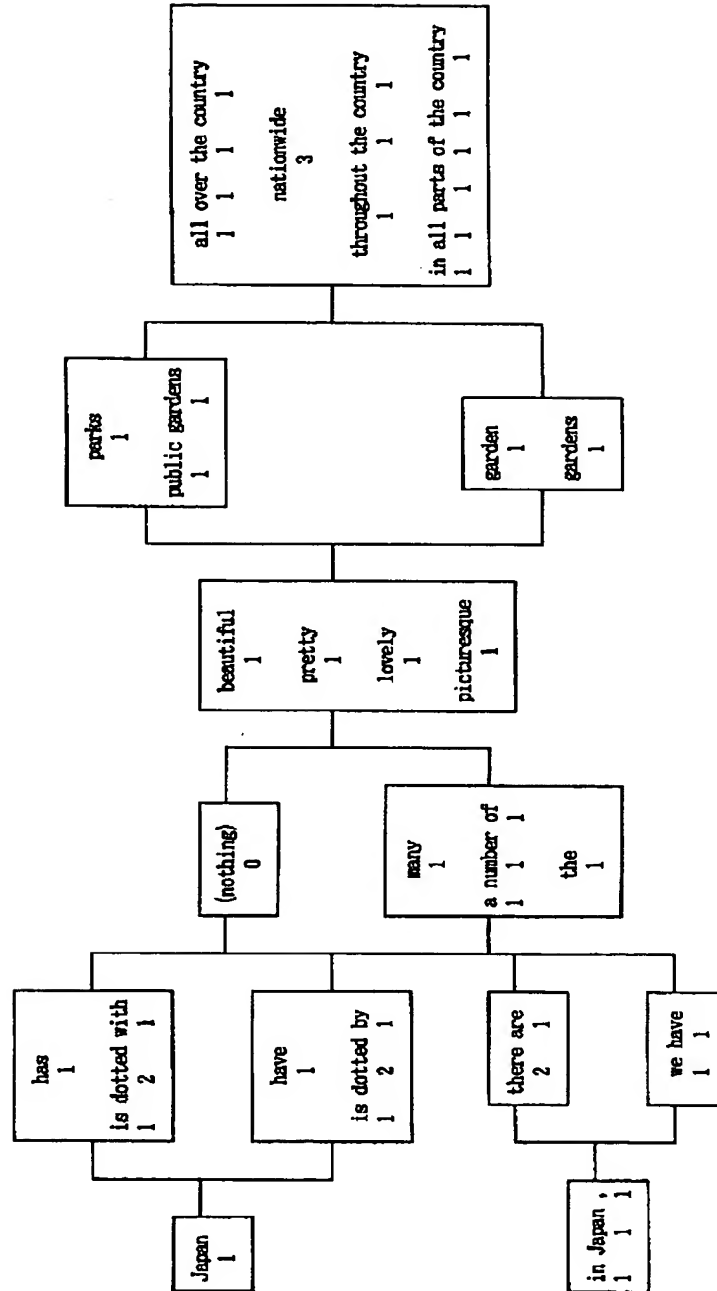
【図1】



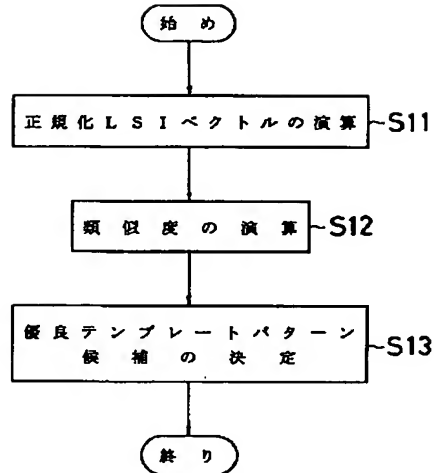
【図3】



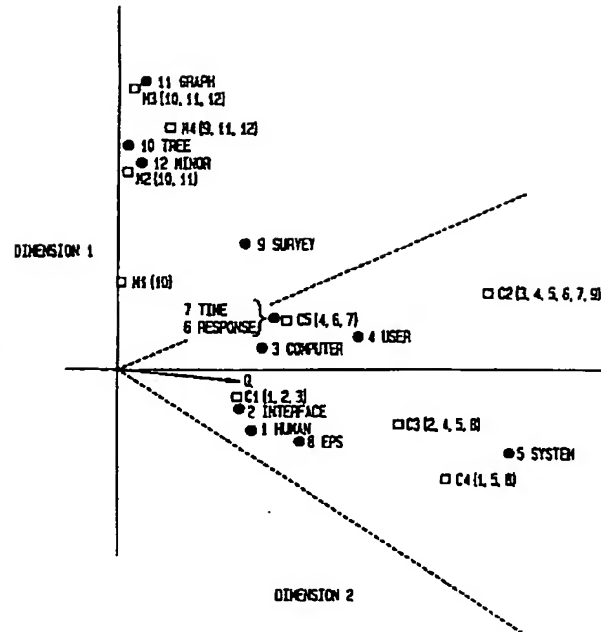
【図2】



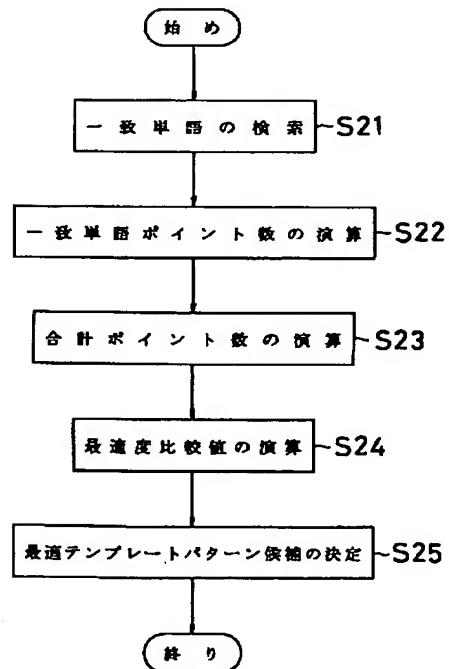
【図4】



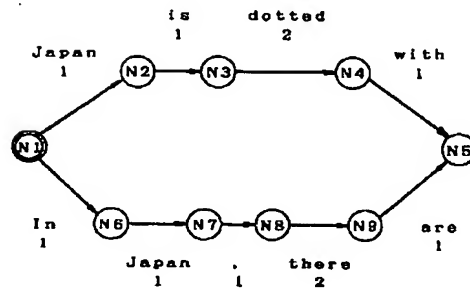
【図5】



【図6】



【図7】



【図8】

	In	Japan	.	are	dotted	with
N1	0	0	0	0	0	0
N2	0	0	0	0	0	0
N3	0	0	0	0	0	0
N4	0	0	0	0	0	0
N5	0	0	0	0	0	0
N6	0	0	0	0	0	0
N7	0	0	0	0	0	0
N8	0	0	0	0	0	0
N9	0	0	0	0	0	0
N5	0	0	0	0	0	0

【図9】

	In	Japan	.	are	dotted	with
N1	0	0	0	0	0	0
N2	0	1	1	1	1	1
N3	0	0	0	0	0	0
N4	0	0	0	0	0	0
N5	0	0	0	0	0	0
N6	0	0	0	0	0	0
N7	0	0	0	0	0	0
N8	0	0	0	0	0	0
N9	0	0	0	0	0	0
N5	0	0	0	0	0	0

【図 10】

In Japan, are dotted with

N 1	0	0	0	0	0	0
N 2	0	1	1	1	1	1
N 3	0	0	0	0	0	0
N 4	0	0	0	0	0	0
N 6	1	1	1	1	1	1
N 7	0	0	0	0	0	0
N 8	0	0	0	0	0	0
N 9	0	0	0	0	0	0
N 5	0	0	0	0	0	0

【図 11】

In Japan, are dotted with

N 1	0	0	0	0	0	0
N 2	0	1	1	1	1	1
N 3	0	2	2	2	2	2
N 4	0	0	0	0	0	0
N 6	1	1	1	1	1	1
N 7	0	0	0	0	0	0
N 8	0	0	0	0	0	0
N 9	0	0	0	0	0	0
N 5	0	0	0	0	0	0

【図 12】

In Japan, are dotted with

N 1	0	0	0	0	0	0
N 2	0	1	1	1	1	1
N 3	0	2	2	2	2	2
N 4	0	4	4	4	4	4
N 6	1	1	1	1	1	1
N 7	0	0	0	0	0	0
N 8	0	0	0	0	0	0
N 9	0	0	0	0	0	0
N 5	0	0	0	0	0	0

【図 13】

In Japan, are dotted with

N 1	0	0	0	0	0	0
N 2	0	1	1	1	1	1
N 3	0	2	2	2	2	2
N 4	0	4	4	4	4	4
N 6	1	1	1	1	1	1
N 7	0	0	0	0	0	0
N 8	0	0	0	0	0	0
N 9	0	0	0	0	0	0
N 5	0	5	5	5	5	5

【図 14】

In Japan, are dotted with

N 1	0	0	0	0	0	0
N 2	0	1	1	1	1	1
N 3	0	2	2	2	2	2
N 4	0	4	4	4	4	4
N 6	1	1	1	1	1	1
N 7	1	2	2	2	2	2
N 8	0	0	0	0	0	0
N 9	0	0	0	0	0	0
N 5	0	5	5	5	5	5

【図 15】

In Japan, are dotted with

N 1	0	0	0	0	0	0
N 2	0	1	1	1	1	1
N 3	0	2	2	2	2	2
N 4	0	4	4	4	4	4
N 6	1	1	1	1	1	1
N 7	1	2	2	2	2	2
N 8	1	3	3	3	3	3
N 9	0	0	0	0	0	0
N 5	0	5	5	5	5	5

【図 16】

In Japan, are dotted with

N 1	0	0	0	0	0	0
N 2	0	1	1	1	1	1
N 3	0	2	2	2	2	2
N 4	0	4	4	4	4	4
N 6	1	1	1	1	1	1
N 7	1	2	2	2	2	2
N 8	1	3	3	3	3	3
N 9	1	5	5	5	5	5
N 5	0	5	5	5	5	5

【図 17】

In Japan, are dotted with

N 1	0	0	0	0	0	0
N 2	0	1	1	1	1	1
N 3	0	2	2	2	2	2
N 4	0	4	4	4	4	4
N 6	1	1	1	1	1	1
N 7	1	2	2	2	2	2
N 8	1	3	3	3	3	3
N 9	1	5	5	5	5	5
N 5	1	6	6	6	6	6